

Methodenkompetenz in der Evaluation¹

*Uwe Schmidt, Johannes Gutenberg-Universität Mainz,
Zentrum für Qualitätssicherung und -entwicklung (ZQ)*

Der Arbeitskreis Aus- und Weiterbildung in der Evaluation gehört zu den jüngeren Arbeitskreisen der DeGEval. Als wesentliche Ziele des Arbeitskreises lassen sich nennen: die Entwicklung von Rahmenempfehlungen für die Aus- und Weiterbildung, die Zusammenfassung bereits bestehender Angebote und die Unterstützung des Aufbaus von entsprechenden Angeboten sowie die unter Qualitätsaspekten selektive Informationsbereitstellung von Aus- und Weiterbildungsangeboten und möglichen Praktikumsplätzen.

Im Hinblick auf die Formulierung von Rahmenempfehlungen ist zunächst die Frage zu stellen, welcher Adressatenkreis angesprochen werden soll. Unter der Prämisse, dass die Empfehlungen nicht nur auf klassische Ausbildungsgänge bspw. in der Form von Masterstudiengängen an Hochschulen begrenzt bleiben sollten, lässt sich der vermutete Bedarf an Aus- und Weiterbildung in der Evaluation im Wesentlichen an drei Gruppen von Evaluatoren orientieren.

Zum einen sind es diejenigen, die eine „sporadische Weiterbildung“ nachfragen. Hierunter sind in der Evaluation Tätige zu verstehen, die Interesse an einzelnen spezifischen Angeboten haben, um ihre praktische Arbeit zu ergänzen. Diese Gruppe ist in der Regel nicht an Formen der Zertifizierung interessiert.

Bei der zweiten Gruppe handelt es sich um Evaluatoren, die eine „sukzessive Weiterbildung“ anstreben. Auch sie nehmen praxisbegleitende Angebote wahr, sind jedoch daran interessiert, einen formalen Abschluss zu erlangen, der auf der Wahrnehmung kumulativer, aufeinander aufbauender Angebote gründen könnte.

Die dritte Gruppe schließlich – so unsere Annahme – besteht aus Personen, die noch keine Evaluationserfahrung haben und einen Abschluss als Evaluator im Rahmen eines Studiums bzw. eines geschlossenen Programms anstreben.

¹ Dieser Beitrag beruht auf dem im Rahmen der Jahrestagung der DeGEval 2001 in Speyer gehaltenen Vortrag vor dem Arbeitskreis Aus- und Weiterbildung in der Evaluation und wurde um solche Aspekte und Fragestellungen erweitert, die in der an den Vortrag anschließenden Diskussion eingebracht wurden.

Unter Zugrundelegung dieser Nachfragegruppen und mit Blick auf die vielfältigen Praxisfelder der Evaluation ist davon auszugehen, dass ein Aus- und Weiterbildungsprogramm in der Evaluation sehr unterschiedlichen Bedürfnissen Rechnung zu tragen hat. Zudem ist anzunehmen, dass auch innerhalb der genannten Gruppen keine homogenen Wissenshintergründe vorhanden sind. So ist es z.B. auch im Hinblick auf ein Studienprogramm wahrscheinlich, dass – zumal dann, wenn es sich um ein Postgraduate-Programm handelt – die Methoden- oder Feldkenntnisse als sehr uneinheitlich einzustufen sind.

Was bedeutet dies für die Entwicklung von Rahmenempfehlungen der Aus- und Weiterbildung? Zunächst ist ein Kanon an Inhalten zu entwickeln, der möglichst umfassend die unterschiedlichen Anforderungen an ein Aus- und Weiterbildungsprogramm abbildet. Damit rückt die Frage in den Mittelpunkt, welche Fertigkeiten grundsätzlich in der Evaluation gefordert sind, ohne bereits die Umsetzung in einzelne Programme oder Programmsegmente zu diskutieren. Darüber hinaus bietet es sich an, die Konzeptionalisierung entlang einzelner Module zu verfolgen, um zum einen den genannten unterschiedlichen Bedürfnissen von Evaluatoren sowie den einzelnen Feldern der Evaluation Rechnung zu tragen. Die Differenzierung von Aus- und Weiterbildungsprogrammen hat zudem den Vorteil, dass auf diese Weise einfacher bereits bestehende Angebote einbezogen werden können. Hierbei ist zu vermuten, dass bereits viele Angebote vorhanden sind, die nicht unter dem Label ‚Evaluation‘ firmieren, gleichzeitig aber sehr gut in die bisherigen Überlegungen zu einem Aus- und Weiterbildungsprogramm in der Evaluation integriert werden könnten.

Die bisherigen Überlegungen des Arbeitskreises führten zu einer Festlegung von vier Modulen. Hierbei handelt es sich um *Methodenkompetenzen* in der Evaluation, *soziale Kompetenzen*, *Feld- und Organisationskenntnisse* sowie um Wissen über die *Geschichte und Theorien der Evaluation*. Entlang dieser Module können weitere Teilbereiche definiert werden.

An dieser Stelle soll zunächst auf die *Methodenkompetenzen* in der Evaluation eingegangen werden. Hierbei lassen sich nach den Vorstellungen des Arbeitskreises zwei wesentliche Bereiche unterscheiden, die grob als wirtschaftswissenschaftliche Methoden und solche der empirischen Sozialforschung definiert werden können. Unter wirtschaftswissenschaftlichen Methoden interessieren all jene Kompetenzen, welche bspw. für die kalkulatorische Planung von Evaluationen nützlich sind. Daneben sind wirtschaftswissenschaftliche Methoden im Sinne von Feldkenntnissen bspw. hilfreich für die Analyse von Prozessen, die auf der Grundlage einer Kosten-Leistungsrechnung gesteuert werden. An dieser Stelle jedoch soll nicht weiter auf wirtschaftswissenschaftliche Kenntnisse als viel mehr auf die Frage nach der für die Evaluation notwendigen empirischen Methodenkompetenz eingegangen werden.

Die in der Evaluation verwendeten empirischen Methoden decken schon aufgrund der unterschiedlichen mit dem Begriff der Evaluation umschriebenen Felder und Zugänge ein weites Spektrum ab. Ohne an dieser Stelle eine grundlegende Auseinandersetzung um den Begriff der Evaluation nachzeichnen zu können, sei auf die große Bandbreite an Bedeutungen hingewiesen, die von einem umfassenden Verständnis der Evaluation als Organisationsentwicklung bis hin zu deskriptiven Datenerhebungen reicht. Betrachtet man unter dieser Perspektive die Diskussion der vergangenen Jahre, so ist weniger zu fragen, was Evaluation ist, sondern viel-

mehr, was nicht mehr unter dem Begriff der Evaluation gefasst wird. In diesem Zusammenhang sind Klärungsversuche hilfreich, die dazu beitragen, nicht jede Form der Befragung oder Datenerhebung unter dem Etikett der Evaluation zu summieren.

Doch auch unter einem engeren Begriffsverständnis dessen, was Evaluation bedeutet, zeigt eine Betrachtung der Evaluationsfelder, dass die Breite und damit verbunden die Methodenvielfalt konstitutiv sind. Längst lässt sich die Evaluation nicht mehr ausschließlich in den Grenzen der Wirkungsforschung von häufig sozialpolitisch orientierten Programmen in Form von „quasi-experimentellen Untersuchungsdesigns“² fassen, sondern sie bezieht sich verstärkt auf die Bewertung ganz unterschiedlicher sozialer Prozesse und Organisationen. Evaluation hat damit in den vergangenen Jahren in einigen Bereichen mehr und mehr eine Entwicklung erfahren, die sie als Steuerungs- und Beratungsinstrument für Fragen der Organisationsentwicklung ausweist. Hiermit verbunden sind Methoden der Datenerhebung und ‚Selbstbeschreibung‘ von Einrichtungen, die in der Anwendung unterschiedlicher Erhebungsinstrumente bestehen. Gleichzeitig aber bleiben die traditionellen, häufig auf quantitativen Analysen basierenden Felder der Evaluation – und insbesondere der Programmevaluation – erhalten.

Dieser fortschreitenden Differenzierung der Evaluationsgegenstände steht der im Rahmen des Arbeitskreises Aus- und Weiterbildung in der DeGEval formulierte Anspruch entgegen, feldübergreifende Methodenkompetenzen zusammenzufassen, die für die Arbeit der Evaluation in unterschiedlichen Bereichen und unter zum Teil gegensätzlichen Prämissen relevant sind. Ein solcher Versuch legt ein Vorgehen nahe, das zunächst um eine Beschreibung derjenigen methodischen Kompetenzen bemüht ist, die als Basiswissen anzusehen sind und der feldspezifischen Explikation bedürfen. Hiermit ist gleichzeitig verbunden, dass solche grundsätzlichen Überlegungen noch nicht auf ihre konkrete Ausgestaltung in Aus- und Weiterbildungsprogrammen diskutiert werden sollen. So ist bspw. mit Blick auf ein Studienprogramm festzuhalten, dass einige der folgend beschriebenen Methodenkompetenzen nicht zwingend Bestandteil eines Aus- und Weiterbildungsprogramms sein müssen, sondern bereits als Kenntnisse vorausgesetzt werden sollten.

Dazu gehört sicherlich eine allgemeine Einführung in die empirische Sozialforschung, die unter anderem grundlegende Erhebungsmethoden, Phasen empirischer Untersuchungen, eine Einführung in wissenschaftstheoretische Grundlagen, die Anwendungsfelder empirischer Sozialforschung und Grundlagen der deskriptiven Statistik vermittelt. Bezieht man die damit beschriebenen Themenfelder auf die Perspektive eines modularisierten Aus- und Weiterbildungsprogramms, so lassen sich auf einer ersten Ebene die folgenden Module hervorheben, für die jeweils exemplarische Aspekte bzw. Problemkontexte dargestellt werden. Zu bedenken ist in diesem Zusammenhang, dass die im Folgenden angeführten Beispiele aus dem Bereich der Evaluation keinesfalls als grundsätzliche bzw. in allen Feldern verbreitete Phänomene zu diskutieren sind. Ganz im Gegenteil finden sich in den Arbeitsfeldern der Evaluation hervorragende Empiriker, die auf einem gänzlich anderen Niveau Methodenprobleme der Evaluation erörtern. Doch dies ist nicht die Regel und die

2 Diekmann, Andreas (1996): Empirische Sozialforschung, 35.

Heterogenität der Evaluationsfelder und der Berufsbiographien von Evaluatoren findet auch in Fragen der Methodenkompetenz ihren Niederschlag.

a) Grundzüge der empirischen Sozialforschung

Hierunter sind zunächst solche Themenbereiche zusammenzufassen, die eine Einführung in die Geschichte der empirischen Sozialforschung, die Planung empirischer Untersuchungen, die wissenschaftstheoretischen Grundlagen und damit in Fragen des Verhältnisses von Theorie und empirischer Forschungspraxis geben. Im Hinblick auf die Evaluation sollte hierbei vor allem den Prozessen der Hypothesenbildung und der Operationalisierung besondere Bedeutung zukommen. Nicht zuletzt die anhaltende Diskussion um Aspekte der Qualität und Leistung von Programmen oder Einrichtungen zeigt, dass in vielen Feldern keinesfalls geklärt ist, unter welchen Kriterien Wirksamkeit oder Qualität gemessen werden. Hierbei entbindet auch nicht der Verweis auf die zumeist sehr allgemein gehaltenen Zielsetzungen, denen Evaluationsgegenstände unterliegen, der Bildung von Dimensionen und Indikatoren sowie der Offenlegung der verwendeten Maßstäbe der Messung von Qualität. Dies scheint um so wichtiger, wenn der Evaluation eine zunehmende Bedeutung als Steuerungsinstrument in Organisationen zukommt. Mit Blick auf die in vielen Bereichen praktizierte und der politischen Logik folgenden Bildung von Kennzahlensystemen für die Verteilung von personellen Ressourcen und Sachmitteln sollte der Evaluation verstärkt die Aufgabe zukommen, wissenschaftlich fundierte Erkenntnisse der Indikatorenbildung einzubringen. Nicht zuletzt deshalb meint Operationalisierung in diesem Zusammenhang auch die Orientierung an Theorien oder zumindest an Erklärungsmodellen für soziale Prozesse. In dieser Hinsicht hat sich die Evaluation zunehmend von der Vorstellung gelöst, theoriegeleitete oder zumindest theoriegestützte Forschung zu betreiben.

b) Datenerhebung

Die Datenerhebung in der Evaluation bedient sich grundsätzlich der klassischen Instrumente der Befragung, der Inhaltsanalyse sowie der Beobachtung, wobei der Befragung nach wie vor eine dominante Stellung zukommt. Nicht für alle, doch für viele Bereiche der Evaluation gilt, dass die Entwicklung von Fragebögen und Frage- sowie Beobachtungsleitfäden eher Plausibilitätsvorstellungen als des Einsatzes geprüfter Instrumente unterliegt. Gerade im Hinblick auf die Entwicklung von Erhebungsinstrumenten ist häufig ein ‚Learning by Doing‘ zu beobachten, das in den vergangenen Jahren in einigen Evaluationsfeldern zu der Produktion einer Vielzahl von Fragebögen ganz unterschiedlicher Qualität geführt hat. Zum Teil sind es grundlegende Kenntnisse über die Fragebogenentwicklung – wie die Verwendung von Fragen mit doppelter Verneinung, sich überschneidende Antwortkategorien, Aufbau des Fragebogens, Verwendung und Bedeutung unterschiedlicher Skalen usw. –, die keine Beachtung finden. Zum Teil fehlt es aber auch an einem kontinu-

ierlichen Erfahrungsaustausch über die Leistungsfähigkeit von Erhebungsinstrumenten und die gewonnenen Ergebnisse im Rahmen von Evaluationen.³

Ebenso deutlich sind die Defizite im Hinblick auf Methoden der Datenerhebung bei dem Einsatz qualitativer Methoden der empirischen Sozialforschung. Gespräche, die im Zusammenhang mit Evaluationen geführt werden, sind – so lässt sich beobachten – in einigen Feldern noch immer eher zufällig als strukturiert und an der Kumulation von Ergebnissen interessiert. In gleicher Weise werden Beobachtungen, soweit sie überhaupt Berücksichtigung in der Evaluation finden, selten als strukturiertes Datenerhebungsinstrument eingesetzt, sondern unterliegen Plausibilitätsvorstellungen, was auch in der Dokumentation der Ergebnisse seinen Niederschlag findet. Die latente Diskussion um die Frage, ob quantitative oder qualitative Verfahren der empirischen Sozialforschung in der Evaluation zum Einsatz kommen sollten, lässt sich meines Erachtens unter dieser Perspektive als nachgeordnetes Problem fassen, zumal grundsätzlich der Einsatz unterschiedlicher Methoden in Evaluationsfeldern allein aufgrund der häufig geringen Fallzahlen als angemessen erscheint.

c) Deskriptive Statistik

Eine Einführung in die deskriptive Statistik sollte im Wesentlichen zu einem sinnvollen Umgang mit Häufigkeitsverteilungen und Mittelwertberechnungen führen. Die in der Evaluation von Programmen, aber auch Organisationen häufig geringen Fallzahlen legen einen vorsichtigen Umgang mit Aussagen über Verteilungen nahe. So zeigt sich bspw., dass nicht selten auf prozentuale Beschreibungen in Fällen zurückgegriffen wird, in denen einzelne Items lediglich mit der zustimmenden Aussage von ein oder zwei Befragten besetzt sind.

Neben diesem Problem geringer Fallzahlen scheint im Hinblick auf den Einsatz deskriptiver Statistik in der Evaluation der Umgang mit Mittelwerten von Bedeutung. Neben den einfachen Verfahren der Berechnung arithmetischer Mittelwerte und des Medians sollte nicht zuletzt bei der Datenerhebung in Form von Skalen der Betrachtung von Abweichungen besondere Bedeutung zukommen. So ist z.B. der partiell zu beobachtende Rückzug auf den bloßen Vergleich von Mittelwerten ohne Berücksichtigung der zugrunde liegenden Verteilungen und etwaiger Polarisierungen von Nennungen selten für eine angemessene Dateninterpretation ausreichend.

d) Grundzüge der schließenden Statistik

Der Einbezug von Grundzügen der schließenden Statistik geschieht zunächst auf dem Hintergrund, dass Evaluatoren über das notwendige Wissen verfügen sollten, um zu beurteilen, welche statistischen Instrumente nicht für den Einsatz in der

3 Hilfreich sind hierbei Mailforen, wie beispielsweise „forum-evaluation“ der Arbeitsstelle für Evaluation pädagogischer Dienstleistungen an der Universität zu Köln (<http://www.uni-koeln.de/ew-fak/wiso/mailling.htm>), die zu einem solchen Austausch beitragen, gleichzeitig aber auch den hohen Bedarf an methodischem Grundwissen in der Evaluation signalisieren.

Evaluation geeignet scheinen. Die Verwendung schließender statistischer Verfahren, wie bspw. Korrelationsanalysen, sind an Kriterien gebunden, die in der Evaluation im Allgemeinen selten erfüllt sind. So handelt es sich in Evaluationen nur in Ausnahmefällen um Zufallsstichproben, die Verallgemeinerungen zulassen. Darüber hinaus sind die Stichproben aufgrund des Modellcharakters, welchen die meisten der evaluierten Projekte oder Programme tragen, so gering, dass Schlüsse über den konkreten Evaluationsgegenstand hinaus nur mit sehr starken Vorbehalten herzustellen sind.

e) Datenverarbeitung, Datenaufbereitung und -interpretation

Ohne an dieser Stelle näher auf die Datenverarbeitung eingehen zu wollen, sollte die Einführung in ein Standard-Statistikprogramm wie SPSS selbstverständlicher Bestandteil einer auf empirische Methoden fokussierten Aus- und Weiterbildung sein.

Betrachtet man die Praxis der Evaluation, so scheint allerdings die Vermittlung von Grundlagen der Datenauswertung und der -interpretation von noch größerer Bedeutung. Aus den eigenen Feldkenntnissen im Bereich der Hochschul- und Bildungsevaluation lässt sich bspw. feststellen, dass häufig bei geringen Fallzahlen mit vielfältigen Differenzierungskriterien gearbeitet wird, was dazu führt, dass Zellen zum Teil mit einem oder wenigen Fällen besetzt sind. Bei der Darstellung der Daten werden solche „Probleme“ nicht selten dadurch umgangen, dass die Interpretation in Form prozentualer Darstellungen erfolgt, was bei kleinen Untersuchungseinheiten zu unsinnigen Ergebnissen führt, die aber ungeachtet dessen als Argumentationshilfen herangezogen werden. Ein weiteres typisches Problem bei der Dateninterpretation ist der Umgang mit Mittelwerten, die häufig auf Grundlage von 5-Punkt- oder 7-Punkt-Likert-Skalen errechnet werden. Hierbei soll weniger die Frage interessieren, ob eine solche Skala überhaupt im Sinne einer Intervallskala interpretiert werden kann, als vielmehr die verkürzte Aufbereitung in Form von Mittelwerten ohne Berücksichtigung von Streuungsmaßen und entsprechenden Polarisierungen der Skalenwerte.

Es ließen sich weitere solcher Detailfragen bei der Dateninterpretation nennen, die eine grundlegende Einführung in Fragen der empirischen Forschung nahe legen. Hierzu gehört meines Erachtens auch eine angemessene Form der Verschriftlichung, die nicht aus einer bloßen Aneinanderreihung von Häufigkeitsverteilungen besteht, sondern auf den Bedeutungsgehalt und die Aussagekraft des Datenmaterials eingeht.

Zusammenfassend lässt sich mit Blick auf die Methodenkompetenz als ein Element der Aus- und Weiterbildung in der Evaluation feststellen, dass es sich hierbei zunächst im Wesentlichen um die Vermittlung von Grundlagen der empirischen Sozialforschung handelt, die allerdings an konkreten Praxisfeldern der Evaluation exemplarisch behandelt werden sollten. Dies schließt ausdrücklich nicht die Einbindung von Veranstaltungen aus, die spezifische methodische Probleme der Evaluation zum Gegenstand haben, und es bedeutet keineswegs, dass die genannten Grundlagen bspw. im Rahmen eines Masterstudiengangs nicht vorauszusetzen wären.